IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of: Barsness et al.	§	Filed: April 22, 2004
Serial No.: 10/829,624	§	Group Art Unit: 2162
Confirmation No.: 7304	§	Examiner: Giovanna B. Colan

For: TECHNIQUES FOR IDENTIFYING MERGEABLE DATA

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF MAILING OR TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450, or facsimile transmitted to the U.S. Patent and Trademark Office to fax number 571-273-8300 to the attention of Examiner Giovanna B. Colan, or electronically transmitted via EFS-Web, on the date shown below:

March 24, 2008 /Jon K. Stewart/
Date Jon K. Stewart

APPEAL BRIEF

Dear Sir:

Applicants submit this Appeal Brief to the Board of Patent Appeals and Interferences on appeal from the decision of the Examiner of Group Art Unit 2162 dated September 18, 2007, finally rejecting claims 1-44. The final rejection of claims 1-44 is appealed. This Appeal Brief is believed to be timely since it is transmitted by the due date of March 24, 2008, as set by the filing of a Notice of Appeal on January 22, 2008.

Please charge the fee of $510.00 for filing this brief to:
Deposit Account No. 09-0465 / ROC920030407US1.

TABLE OF CONTENTS

Real Party in Interest

The present application has been assigned to International Business Machines Corporation, Armonk, New York.

Related Appeals and Interferences

Applicant asserts that no other appeals or interferences are known to the Applicant, the Applicant's legal representative, or assignee which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

Status of Claims

Claims 1-44 are pending in the application. Claims 1-44 were originally presented in the application. Claims 1-44 stand finally rejected as discussed below. The final rejections of claims 1-44 are appealed. The pending claims are shown in the attached Claims Appendix.

Status of Amendments

All claim amendments have been entered by the Examiner. No amendments to the claims were proposed after the final rejection.

Summary of Claimed Subject Matter

Claimed embodiments include methods (*see e.g.,* claims 1-16, 17-21), computer programs stored on computer readable storage media (*see e.g.*, claims 22-37, 38-42) and computer systems (*see e.g.,* claims 43-44) directed for identifying mergeable data in a data processing system and, more particularly, for identifying correlated columns from one or more database tables.

A. CLAIM 1 – INDEPENDENT

Claim 1 is directed to a computer-implemented method for identifying correlated columns from database tables. *See Application*, 1:5-7, 3:14-16, 3:18-27, 7:6-11. As claimed, this method includes determining correlation attributes for a first column and a second column from one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column. *See Application*, 12:1-14, Figure 1, 130, 140, Figure 3, 300, 7:13-20, 14:8-18, Figure 2, 230, 240. This method also includes comparing the correlation attributes from the first and second column. *See Application*, 7:20-22, Figure 1, 130, 140 12:16-31, Figure 4, 400. And also includes identifying similarities between the first and second column on the basis of the comparison. *See Application*, 7:20-22, 13:3-17, 14:20-25, Figure 2, 250. On the basis of the identified similarities, this method includes determining whether the first and second column are correlated. *See Application*, 7:22-23, 14:27-31, Figure 2, 260, 15:1-4. Upon determining the first and second columns are correlated, this method includes merging the first and second columns to create a third column that contains each data value stored in the first and second columns, *see Application*, 8:1-5, 15:6-10, Figure 2, 270, and storing the third column in the database *see Application*, 8:1-5, 15:6-10, Figure 2, 270.

B. CLAIMS 2 AND 23 – DEPENDENT

Claim 2 depends from claim 1 and specifies that the step of identifying the similarities recited by claim 1 includes determining a correlation value indicating a

degree of correlation between the first and the second column. See Application 13:3-17, 18:6-26. Claim 2 also adds a step of determining whether the correlation value exceeds a predetermined threshold. *See Application*, 13:13-17, 21:28-32, 22:5-9, 22:15-19, 22:24-27. Claim 23 recites the same limitation relative to independent claim 22.

C. CLAIMS 9 AND 30 – DEPENDENT

Claim 9 specifies the method of claim 1, with a further step of determining, from the one or more database tables, statistical parameters associated with each of the columns. *See Application*, 21:22-26, Figures 6A, 6B, 600. Claim 9 further specifies that the correlation attributes are determined on the basis of the determined statistical parameters. *See Application*, 21:22-26, Figures 6A, 6B, 600. Claim 30 recites the same limitation relative to independent claim 22.

D. CLAIM 11 AND 34 – DEPENDENT

Claim 11 specifies the method of claim 1, with a further step of determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column. *See Application*, 23:18-28, Figure 7 700, Figure 8, et seq. Claim 11 further specifies that the correlation attributes are determined on the basis of the determined ontological properties. *See Application*, 23:18-28, Figure 7 700, Figure 8, *et seq*. Claim 34 recites the same limitation relative to independent claim 22.

E. CLAIM 14 AND 35 DEPENDENT

Claim 14 specifies the method of claim 1, with a further step of determining, from the one or more database tables, measurement units associated with each column. *See Application*, 26:4-20, Figure 9, 900 and accompanying description, Table I, p. 27, *et seq.* Claim 14 further specifies that the correlation attributes are determined on the basis of the determined measurement units. *See Application*, 23:18-28, Figure 7 700, Figure 8, *et seq*. Claim 35 recites the same limitation relative to independent claim 22.

F. CLAIM 17 – INDEPENDENT

Claim 17 is directed to a computer-implemented method for identifying correlated columns from database tables. *See Application*, 1:5-7, 3:29-30 – 4:1-5, 3:18-27, 7:6-11. As claimed, this method includes determining metadata for at least two columns from one or more database tables, the metadata describing characteristics of each column. *See Application*, 12:22-31, 12:1-14, 8:7-20, 39:6-11, Figure 15, 1520. As claimed, this method also includes analyzing content from the at least two columns from the one or more database tables. *See Application*, 7:26-29, 8:21-31 – 9:1-12, 39:13-19, Figure 15, 1540. This method also includes determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content, *see Application*, 39:21-31, 40:1-3, and storing the value representing the degree of correlation in the database *see Application*, 39:21-31, 40:1-3.

G. CLAIM 22 – INDEPENDENT

Claim 22 is directed to a computer readable storage medium containing a program which, when executed, performs a process for identifying correlated columns from database tables. *See Application*, 1:5-7, 4:7-16, 3:18-27, 7:6-11, 9:15-27. As claimed, the process includes determining correlation attributes for a first column and a second column from one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column. *See Application*, 12:1-14, Figure 1, 130, 140, Figure 3, 300, 7:13-20, 14:8-18, Figure 2, 230, 240. This process also includes comparing the correlation attributes from the first and second column. *See Application*, 7:20-22, Figure 1, 130, 140 12:16-31, Figure 4, 400. And also includes identifying similarities between the first and second column on the basis of the comparison. *See Application*, 7:20-22, 13:3-17, 14:20-25, Figure 2, 250. On the basis of the identified similarities, this process also includes determining whether the first and second column are correlated. *See Application*, 7:22-23, 14:27-31, Figure 2, 260, 15:1-4. Upon determining the first and second columns are correlated, this process also includes merging the first and second columns to create a third column that contains each data value stored in the first and second columns, *see Application*,

8:1-5, 15:6-10, Figure 2, 270, and storing the third column in the database *see Application*, 8:1-5, 15:6-10, Figure 2, 270.

H. CLAIM 38 – INDEPENDENT

Claim 38 is directed to a computer readable storage medium containing a program which, when executed, performs a process for identifying correlated columns from database tables. *See Application*, 1:5-7, 4:18-25, 3:18-27, 7:6-11, 9:15-27. As claimed, this process includes determining metadata for at least two columns from one or more database tables, the metadata describing characteristics of each column. *See Application*, 12:22-31, 12:1-14, 8:7-20, 39:6-11, Figure 15, 1520. This process also includes analyzing content from the at least two columns from the one or more database tables. *See Application*, 7:26-29, 8:21-31 – 9:1-12, 39:13-19, Figure 15, 1540. This process also includes determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content, *see Application*, 39:21-31, 40:1-3, and storing the value representing the degree of correlation in the database *see Application*, 39:21-31, 40:1-3.

I. CLAIM 43 – INDEPENDENT

Claim 43 is directed to a data processing system. *See Application*, 1:5-7, 4:27-31 – 5:1-7, 3:18-27, 7:6-11, 10:11-17. As claimed, the system includes a processor (see application 10:11-17) and at least one database having one or more database tables. *See Application*, 11:19-30, 12:1-15, Figure 1, 110, 120. As claimed, the system also includes a correlation manager for identifying correlated columns from the one or more database tables. *See Application*, 12:16-31, 13:1-17, Figure 1, 150. As claimed, the correlation manager is configured for determining correlation attributes for a first column and a second column from the one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column. *See Application*, 12:1-14, Figure 1, 130, 140, Figure 3, 300, 7:13-20, 14:8-18, Figure 2, 230, 240. The correlation manager is further configured for comparing the correlation attributes from the first and second column, see *Application*, 7:20-22, Figure 1, 130, 140 12:16-31, Figure 4, 400, and identifying similarities between the first and

second column on the basis of the comparison. *See Application*, 7:20-22, 13:3-17, 14:20-25, Figure 2, 250. The correlation manager is further configured for, on the basis of the identified similarities, determining whether the first and second column are correlated. *See Application*, 7:22-23, 14:27-31, Figure 2, 260, 15:1-4. Upon determining the first and second columns are correlated, the correlation manager is configured to merge the first and second columns to create a third column that contains each data value stored in the first and second columns, see *Application*, 8:1-5, 15:6-10, Figure 2, 270, and storing the third column in the database *see Application*, 8:1-5, 15:6-10, Figure 2, 270.

J. CLAIM 44 – INDEPENDENT

Claim 44 is directed to a data processing system. As claimed, the system includes a processor (see application 10:11-17) and at least one database having one or more database tables. *See Application*, 11:19-30, 12:1-15, Figure 1, 110, 120. As claimed, the system also includes a correlation manager for identifying correlated columns from the one or more database tables. *See Application*, 12:16-31, 13:1-17, Figure 1, 150. As claimed, the correlation manager is configured for determining metadata for at least two columns from the one or more database tables, the metadata describing characteristics of each column. *See Application*, 12:22-31, 12:1-14, 8:7-20, 39:6-11, Figure 15, 1520. As recited by claim 44, the correlation manager is further configured for analyzing content from the at least two columns from the one or more database tables. *See Application*, 7:26-29, 8:21-31 – 9:1-12, 39:13-19, Figure 15, 1540. As recited by claim 44, the correlation manager is further configured for determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content, s*ee Application*, 39:21-31, 40:1-3, and storing the value representing the degree of correlation in the database *see Application*, 39:21-31, 40:1-3.

Grounds of Rejection to be Reviewed on Appeal

Rejection of claims 1-44 under 35 U.S.C. § 103(a) as being unpatentable over *Sandler et al.*, U.S. Patent Application Publication No. 2003/0217033 A1 (hereinafter *Sandler*) in view of *Kaufman et al.*, U.S. Patent Application Publication No. 2004/0073565 A1 (hereinafter *Kaufman*).

ARGUMENTS

Sandler in view of *Kaufman* do not render Claims 1-44 Obvious under 35 U.S.C. § 103(a)

The Applicable Law

The Examiner bears the initial burden of establishing a *prima facie* case of obviousness. *See* MPEP § 2142. To establish a *prima facie* case of obviousness three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one ordinary skill in the art to modify the reference or to combine the reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. *See* MPEP § 2143. The present rejection fails to establish at least the third criteria.

Regarding claims 1, 22, and 43:

Applicants submit that *Sandler* does not disclose a "method for identifying correlated columns from database tables" that includes "determining correlation attributes for a first column and a second column from one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column," as recited by claim 1. Claims 22 and 43 recite a similar limitation. Nevertheless, the Examiner suggests that *Sandler* discloses this limitation as follows:

> Fig. 18A, items 1806, and 1804, Page 17, [0235], lines 7 - 12; "...all of the values in field K1 1804 that have the same values in field F1 1806..." wherein the step of mapping which includes all of the values in field K1 1804 that have the same values in field F1 1806 corresponds to the step of determining the correlation attributes as claimed; wherein values F1 corresponds to the first column claimed; and wherein values in K1 corresponds to the second column claimed;

Advisory Action, continuation sheet. This argument simply repeats, virtually word-for-word, the Examiner's position taken in a Final Office Action. (*See Final Office Action*, p. 4.) In fact however, the values in "field K1 1804 that have the same values in field F1 1806" do not teach or suggest "correlation attributes" as suggested by the Examiner.

The values in the example field "K1" of *Sandler* are not correlation attributes of the values in the field "F1" at all; instead, the field values are simply data values of two columns in the example "T1" table. The material cited by the Examiner describes an example of an "aggregation" operation disclosed in *Sandler* used to add certain values in a table together. As disclosed in *Sandler*:

> [s]uch aggregation operations are used to represent many-to-one relations, and occur only after the table rule has been applied, to convert a combined table (which results from application of various fuse, link, and loop operations) into the target table.

Sandler, ¶ 234. The specific Example cited by the Examiner includes a table from Figure 18A with the following values:

TableT1 (Figure 18, 1800)	
F1	K1
A	1
A	2
B	3
K	4
B	5

In this example, as part of the "aggregation operation" the repeated "A" values of "1" and "2" are summed to create an entry of "A" with a value of "3. Similarly, the repeated "B" values of "3" and "5" are summed to create an entry of "B" with a value of "8". Thus, the "aggregation operation" results in the following table:

Table Target (Figure 18, 1802)	
F1	K1
A	3
B	8
K	4

No part of this process of simply summing up the numerical values in the K1 column, based on repeating values in the F1 column, discloses the claimed step of "determining correlation attributes for a first column and a second column from one or more database tables, as recited by claims 1, 22 and 43.

Further, nothing in this material discloses the claimed steps of "identifying similarities between the first and second column on the basis of the comparison," and

"on the basis of the identified similarities, determining whether the first and second columns are correlated." Instead, this material describes an "aggregation operation" used to process certain entries in a table that result from "loop, fuse, and link" operations.

The Examiner then relies on this same material to argue that *Sandler* discloses the claimed steps of "identifying similarities between the first and second column on the basis of the comparison," and "on the basis of the identified similarities, determining whether the first and second column are correlated," and "upon determining the first and second columns are correlated, merging the first and second columns to create a third column that contains each data value stored in the first and second columns," as recited by claim 1. Specifically, the Examiner suggests:

> [*Sandler* discloses] upon determining the first and second columns are correlated, merging the first and second columns to create a third column (Fig. 18A, Page 17, [0235], lines 8 -15, "To perform this mapping, all of the values in field K1 1804 **that have the same values** in field F1 1806 must be combined ... "; wherein "the same values" corresponds to the identified similarities claimed; Sandler).

Final Office Action, p. 5. Clearly, however, the only columns present in the example from *Sandler* are the F1 and K1 columns. There is simply no third column present. Instead, the K1 column has an initial state (without aggregated values) and a target state (with aggregated values). Accordingly, Applicants respectfully request that the rejection of claims 1, 22, and 43 be vacated.

Furthermore, the deficiency with the present rejection is readily apparent when viewed in light of the limitations recited by dependent claims 2, 6, 9, 11, 14 and 23. Each of these claims further characterizes the "correlation attributes" recited by the corresponding independent claim.

Regarding claims 2 and 23:

> Claim 2 depends from claim 1 and recites:
>
> The method of claim 1, wherein identifying the similarities comprises:
> determining a correlation value indicating a degree of correlation between the first and the second column; and

> determining whether the correlation value exceeds a predetermined threshold.

Claim 23 recites a similar limitation relative to claim 22. In rejecting these claims, the Examiner suggests that *Sandler* discloses:

> [a step of] determining whether the correlation value exceeds a predetermined threshold (Page 2, [0018], lines 6 - 9; is above a predetermined threshold; *Sandler*).

Final Office Action, p. 6. This passage provides:

> In general, in another aspect, the invention relates to a method for external checkpointing. The method includes initially communicating a data table and a log comprising entries of data table transactions to a subscriber; and communicating additional log entries to the subscriber when they are received. The method includes determining that the number of log entries is above a predetermined threshold, applying the log entries to the data table, and communicating the updated data table to the subscriber.

Sandler, ¶ 18. This description of "method for external checkpointing" that includes "determining whether that number of log entries is above a predetermined threshold" has no relationship whatsoever to the claimed subject matter. As is well known, "checkpointing" refers to a synchronization point between data files and log files. The cited passage describes transaction processing that includes executing log entries once the number of log entries reaches "a predetermined threshold." Clearly, the use of "a predetermined threshold" to determine when the number of log entries is above that threshold has nothing to do with the subject matter of "determining a correlation value indicating a degree of correlation between the first and the second column" and "determining whether the correlation value exceeds a predetermined threshold," as claimed. The former is related to a threshold for a number of entries in a log file, whereas the claimed limitation is directed to a threshold for a determined degree of correlation between the first column and second column recited in the independent claims. Furthermore, the passage describing a threshold for the number of entries in a log file has no relationship whatsoever with the example "T1" table which the Examiner suggests discloses the first column and the second column in the first place.

Accordingly, Applicants respectfully request that the rejection of claims 2 and 23 be vacated.

Regarding claims 9, 11, 14, 30, 34, and 35:

Claims 9, 11, and 14 each depend from claim 1 and further characterize the correlation attributes. Claims 9, 11, and 14 specify that the "correlation attributes" may be determined on the basis of: (i) statistical parameters associated with the column, (claim 9), (ii) ontological properties describing cognitive qualities associated with the column, (claim 11), and (iii) measurement units associated with the column. Claims 30, 34, and 35 recite similar limitations, respectively. The following table compares the passages cited by the examiner against the recited limitations of claims 9, 11, and 14.

Claim#	Claim Limitation	Cited Passages
9, 14	The method of claim 1, further comprising: determining, from the one or more database tables, statistical parameters associated with each of the columns; and **wherein the correlation attributes are determined on the basis of the determined statistical parameters**. (claim 9) The method of claim 1, further comprising: determining, from the one or more database tables, measurement units associated with each column; and **wherein the correlation attributes are determined on the basis of the determined measurement units**. (claim 14)	In one embodiment, the system maintains range indices for each key field column and index column stored in the database. For each fixed-size "chunk" of a column (or index), the range indices contain maximum and minimum values of the data in that chunk. This information can be used to increase the efficiency of certain operations, such as table joins, searches, or identifying minimum or maximum values, without requiring significant additional storage relative to the size of a column. *Sandler*, ¶ 59.
11	The method of claim 1, further comprising: determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column; and **wherein the correlation attributes are determined on the basis of the determined ontological properties.** (Claim 11)	Next, S(G) is exploded into a single table ES(G) (Step 1004). Duplicate fields in ES(G) are discarded to form table UES(G) and a synonym table Z(G) is constructed from ES(G) (Step 1008). UES(G) is partitioned into PUES(G) (Step 1012), and G is constructed by applying the computational rules to PUES(G) and Z(G) (Step 1016). *Sandler*, ¶ 0110.

First regarding claims 9 and 14, the cited passages describe how a database column may include an index. As is well known, a database index is a data structure that improves the speed of database operations by, as the name implies, indexing what values are present in a table. For example, consider a column storing last names, in

four "chunks." In such a case, the index for the "chunks" may partition what records are in an index alphabetically (e.g., A-G, H-M, N-S, and T-Z). This allows a search operation for a given last name be evaluated using only one of the chunks. At the same time, nothing in this generic description of an index discloses the claimed step of determining correlation attributes **between two columns** using statistical information. Similarly, nothing in this description discloses the claimed step of determining correlation attributes **between two columns** using measurement units associated with each column. That is, nothing in this description of a database index discloses anything whatsoever to do with "measurement units." For example, consider the following example from Applicants' specification.

> [0036] For instance, some values may be numeric and represent milligrams, whereas other values may be equivalent but in different units such as "ounces" or "grains". However, both values relate to data directed to masses or weights.
> [0086] For example, the units label for the first column may be "kilograms" while the units label for the second column may be "kg". Nevertheless, these labels may be recognized as being interchangeable, and therefore correlated.

Plainly, the claimed "measurement units associated with each column" are not disclosed by a passage describing a database index used to specify what portions of a column are in a given "chunk" of a table.

Regarding claim 11, the passage cited by the Examiner appears to have nothing whatsoever to do with "determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column." Instead the cited passage describes processing a table named "S(G)," which is described as "processing of each g-table G occurs in six steps. First, all of the input tables I(G) are combined into a **single table S(G)** (Step 1000)." *Sandler*, ¶ 0110. The "g-table" itself is described as follows:

> Intermediary, derived tables are referred to as g-tables or gammas. The fields in a g-table are derived from the rules and one or more tables that the user has previously specified.

The passage describes exploding the S(G) table into ES(G), discarding duplicate elements from ES(G) to form UES(G) and constructing another table Z(G) from ES(G). The passage also describe partitioning the UES(G) table to form PUES(G).

Applicants respectfully submit that none of this comes even close to disclosing the claimed "determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column", as far as the rejection can be understood.

Furthermore, in rejecting claims 9, 11, 14, 30, 34, and 35, the Examiner cites to various portions of *Sandler* unrelated to the "aggregation operation" cited in the rejection of the underlying independent claim. Applicants submit that the isolated fragments cited by the Examiner fail to disclose the claimed characterizations of the "correlation attributes," of the independent claims. Accordingly, Applicants respectfully request that the rejection of claims 9, 11, 14, 30, 34, and 35 be vacated.

Regarding claims 3-6, 7-8, 10, 12-13, 15-16, 24-27, 28-29, 31-33, and 36-37:

Each of claims 3-5, 7-8, 10, 12-13, 15-16, 24-26, 28-29, 31-33, and 36-37 depend from one of claims 1 or 22. Accordingly, for all the reasons given above regarding independent claims 1 and 22, Applicant submits that these dependent claims are allowable and respectfully requests that the rejection of these dependent claims be vacated.

Regarding claims 17, 38, and 44:

Sandler does not teach or suggest a "method for identifying correlated columns from database tables" that includes a step of "determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content." Claims 38 and 44 recite a similar limitation. In rejecting these claims the Examiner again turns to the "aggregation operation" disclosed in *Sandler* used to combine values in a column of a database table. However, as well demonstrated above, the process of "combining the same values in fields" in no way discloses the step of "determining the degree of correlation" between a first database column and a second database column. Instead, the material describes adding numbers in one field of records that share a value in another field of the records. No "degree of correlation" is determined, calculated, or otherwise suggested by the "aggregation operation" disclosed in *Sandler*. Accordingly, Applicant submits that these independent claims are

allowable and respectfully requests that the rejection of claims 17, 38, and 44 be vacated.

Regarding claims 18-21 and 39-42:

Each of claims 18-21 and 39-42 depend from one of claims 17 or 38. Accordingly, for all the reasons given above regarding independent claims 17 and 38, Applicant submits that these dependent claims are allowable and respectfully requests that the rejection of these dependent claims be vacated.

CONCLUSION

The Examiner errs in finding that claims 1-44 are unpatentable over *Sandler* in view of *Kaufman* under 35 U.S.C. § 103(a).

Withdrawal of the rejections and allowance of all claims is respectfully requested.

Respectfully submitted, and

S-signed pursuant to 37 CFR 1.4,

/Gero G. McClellan, Reg. No. 44,227/

Gero G. McClellan
Registration No. 44,227
Patterson & Sheridan, L.L.P.
3040 Post Oak Blvd. Suite 1500
Houston, TX 77056
Telephone: (713) 623-4844
Facsimile: (713) 623-4846
Attorney for Appellant(s)

CLAIMS APPENDIX

1. (Previously Presented) A computer-implemented method for identifying correlated columns from database tables, comprising:

determining correlation attributes for a first column and a second column from one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column;

comparing the correlation attributes from the first and second column;

identifying similarities between the first and second column on the basis of the comparison;

on the basis of the identified similarities, determining whether the first and second column are correlated;

upon determining the first and second columns are correlated, merging the first and second columns to create a third column that contains each data value stored in the first and second columns; and

storing the third column in the database.

2. (Original) The method of claim 1, wherein identifying the similarities comprises:

determining a correlation value indicating a degree of correlation between the first and the second column; and

determining whether the correlation value exceeds a predetermined threshold.

3. (Original) The method of claim 1, further comprising:

if it is determined that the first and second column are correlated, displaying an indication to a user that the first and second column can be merged; and

in response to user input, merging the first and second column into a single column.

4. (Original) The method of claim 1, wherein the first column is a column of a first database table and the second column is a column of a second database table, the method further comprising:

determining correlation attributes for N columns from the first database table and M columns from the second database table, where N and M are integers;

comparing the correlation attributes from each of the N columns with the correlation attributes from each of the M columns to identify similarities between the N and M columns; and

on the basis of the identified similarities, determining whether one or more of the N and M columns are correlated.

5. (Original) The method of claim 4, further comprising merging each of the one or more of the N and M columns determined to be correlated.

6. (Original) The method of claim 1, further comprising:

determining, from the one or more database tables, metadata describing characteristics of each column; and

wherein the correlation attributes are determined on the basis of the determined metadata.

7. (Original) The method of claim 6, wherein the determined metadata describes for each column an attribute of a data value in the column.

8. (Original) The method of claim 6, wherein the determined metadata describes for each column at least one of:

(i) a label;

(ii) a comment;

(iii) a constraint;

(iv) a trigger;

(v) a name;

(vi) a data type; and

(vii) a column length.

9. (Original) The method of claim 1, further comprising:

determining, from the one or more database tables, statistical parameters associated with each of the columns; and

wherein the correlation attributes are determined on the basis of the determined statistical parameters.

10. (Original) The method of claim 9, wherein the determined statistical parameters describe for each column at least one of:

(i) a minimum value;

(ii) a maximum value;

(iii) an average value; and

(iv) a range of values.

11. (Original) The method of claim 1, further comprising:

determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column; and

wherein the correlation attributes are determined on the basis of the determined ontological properties.

12. (Original) The method of claim 11, wherein the determined ontological properties describe for each column at least one of:

(i) a synonym;

(ii) a parent node; and

(iii) an ancestor node.

13. (Original) The method of claim 11, further comprising:

determining, from the one or more database tables, metadata describing the ontological properties.

14. (Original) The method of claim 1, further comprising:

determining, from the one or more database tables, measurement units associated with each column; and

wherein the correlation attributes are determined on the basis of the determined measurement units.

15. (Original) The method of claim 14, further comprising:

determining, from the one or more database tables, metadata describing the measurement units.

16. (Original) The method of claim 14, wherein identifying the similarities comprises:

determining whether the first and second column are associated with similar measurement units.

17. (Previously Presented) A computer-implemented method for identifying correlated columns from database tables, comprising:

determining metadata for at least two columns from one or more database tables, the metadata describing characteristics of each column;

analyzing content from the at least two columns from the one or more database tables;

determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content; and

storing the value representing the degree of correlation in the database.

18. (Original) The method of claim 17, wherein determining the degree of correlation comprises:

assigning a first correlation value to the determined metadata;

assigning a second correlation value to the analyzed content, wherein the first and second correlation values are different; and

calculating a total correlation value on the basis of the first and second correlation values.

19. (Original) The method of claim 18, further comprising:

merging the at least two columns if the total correlation value exceeds a predetermined threshold value.

20. (Original) The method of claim 17, wherein analyzing the content comprises determining statistical parameters from the content of each column.

21. (Original) The method of claim 17, further comprising:

merging the first and the at least one second column if it is determined that the first and at least one second column are correlated.

22. (Previously Presented) A computer readable storage medium containing a program which, when executed, performs a process for identifying correlated columns from database tables, the process comprising:

determining correlation attributes for a first column and a second column from one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column;

comparing the correlation attributes from the first and second column;

identifying similarities between the first and second column on the basis of the comparison;

on the basis of the identified similarities, determining whether the first and second column are correlated;

upon determining the first and second columns are correlated, merging the first and second columns to create a third column that contains each data value stored in the first and second columns; and

storing the third column in the database.

23. (Original) The computer readable medium of claim 22, wherein identifying the similarities comprises:

determining a correlation value indicating a degree of correlation between the first and the second column; and

determining whether the correlation value exceeds a predetermined threshold.

24. (Original) The computer readable medium of claim 22, wherein the process further comprises:

if it is determined that the first and second column are correlated, displaying an indication to a user that the first and second column can be merged; and

in response to user input, merging the first and second column into a single column.

25. (Original) The computer readable medium of claim 22, wherein the first column is a column of a first database table and the second column is a column of a second database table, the process further comprising:

determining correlation attributes for N columns from the first database table and M columns from the second database table, where N and M are integers;

comparing the correlation attributes from each of the N columns with the correlation attributes from each of the M columns to identify similarities between the N and M columns; and

on the basis of the identified similarities, determining whether one or more of the N and M columns are correlated.

26. (Original) The computer readable medium of claim 25, wherein the process further comprises:

merging each of the one or more of the N and M columns determined to be correlated.

27. (Original) The computer readable medium of claim 22, wherein the process further comprises:

determining, from the one or more database tables, metadata describing characteristics of each column; and

wherein the correlation attributes are determined on the basis of the determined metadata.

28. (Original) The computer readable medium of claim 27,wherein the determined metadata describes for each column an attribute of a data value in the column.

29. (Original) The computer readable medium of claim 27, wherein the determined metadata describes for each column at least one of:

(i) a label;

(ii) a comment;

(iii) a constraint;

(iv) a trigger;

(v) a name;

(vi) a data type; and

(vii) a column length.

30. (Original) The computer readable medium of claim 22, wherein the process further comprises:

determining, from the one or more database tables, statistical parameters associated with each of the columns; and

wherein the correlation attributes are determined on the basis of the determined statistical parameters.

31. (Original) The computer readable medium of claim 30, wherein the determined statistical parameters describe for each column at least one of:

(i) a minimum value;

(ii) a maximum value;

(iii) an average value; and

(iv) a range of values.

32. (Original) The computer readable medium of claim 22, wherein the process further comprises:

determining, from the one or more database tables, ontological properties describing cognitive qualities associated with each column; and

wherein the correlation attributes are determined on the basis of the determined ontological properties.

33. (Original) The computer readable medium of claim 32, wherein the determined ontological properties describe for each column at least one of:

(i) a synonym;

(ii) a parent node; and

(iii) an ancestor node.

34. (Original) The computer readable medium of claim 32, wherein the process further comprises:

determining, from the one or more database tables, metadata describing the ontological properties.

35. (Original) The computer readable medium of claim 22, wherein the process further comprises:

determining, from the one or more database tables, measurement units associated with each column; and

wherein the correlation attributes are determined on the basis of the determined measurement units.

36. (Original) The computer readable medium of claim 35, wherein the process further comprises:

determining, from the one or more database tables, metadata describing the measurement units.

37. (Original) The computer readable medium of claim 35, wherein identifying the similarities comprises:

determining whether the first and second column are associated with similar measurement units.

38. (Previously Presented) A computer readable storage medium containing a program which, when executed, performs a process for identifying correlated columns from database tables, the process comprising:

determining metadata for at least two columns from one or more database tables, the metadata describing characteristics of each column;

analyzing content from the at least two columns from the one or more database tables; and

determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content; and

storing the value representing the degree of correlation in the database.

39. (Original) The computer readable medium of claim 38, wherein determining the degree of correlation comprises:

assigning a first correlation value to the determined metadata;

assigning a second correlation value to the analyzed content, wherein the first and second correlation values are different; and

calculating a total correlation value on the basis of the first and second correlation values.

40. (Original) The computer readable medium of claim 39, wherein the process further comprises:

merging the at least two columns if the total correlation value exceeds a predetermined threshold value.

41. (Original) The computer readable medium of claim 38, wherein analyzing the content comprises:

determining statistical parameters from the content of each column.

42. (Original) The computer readable medium of claim 38, wherein the process further comprises:

merging the first and the at least one second column if it is determined that the first and at least one second column are correlated.

43. (Previously Presented) A data processing system, comprising:

a processor;

at least one database having one or more database tables; and

a correlation manager for identifying correlated columns from the one or more database tables, the correlation manager which, when executed by the processor, is configured for:

determining correlation attributes for a first column and a second column from the one or more database tables, the correlation attributes describing for each column at least one of the column and content of the column;

comparing the correlation attributes from the first and second column;

identifying similarities between the first and second column on the basis of the comparison;

on the basis of the identified similarities, determining whether the first and second column are correlated;

upon determining the first and second columns are correlated, merging the first and second columns to create a third column that contains each data value stored in the first and second columns; and

storing the third column in the database.

44. (Previously Presented) A data processing system, comprising:

a processor;

at least one database having one or more database tables; and

a correlation manager for identifying correlated columns from the one or more database tables, the correlation manager which, when executed by the processor, is configured for:

determining metadata for at least two columns from the one or more database tables, the metadata describing characteristics of each column;

analyzing content from the at least two columns from the one or more database tables; and

determining a degree of correlation between the at least two columns using the determined metadata and the analyzed content; and

storing the value representing the degree of correlation in the database.

EVIDENCE APPENDIX

None.

RELATED PROCEEDINGS APPENDIX

None.